Optimization of k-Space Trajectories for Compressed Sensing by Bayesian Experimental Design

Matthias Seeger[1], Hannes Nickisch[2], Rolf Pohmann[2], Bernhard Schölkopf[2]
[1] Department of Computer Science, Saarland University, Saarbrücken
[2] Max Planck Institute for Biological Cybernetics, Tübingen

Running head: k-Space Optimization by Bayesian Experimental Design

Corresponding author: Matthias Seeger, mseeger@mmci.uni-saarland.de,
Max Planck Institute for Informatics, Campus E1.4, Room 116,
66123 Saarbrücken, Germany, Tel. +49 681 9325452

Word count: Main text: ≈4880 words
Appendix: ≈1100 words
Captions: ≈520 words
Number of references: 38
Number of figures: 7

Abstract

The optimization of k-space sampling for nonlinear sparse MRI reconstruction is phrased as Bayesian experimental design problem. Bayesian inference is approximated by a novel relaxation to standard signal processing primitives, resulting in an efficient optimization algorithm for Cartesian and spiral trajectories. On clinical resolution brain image data from a Siemens 3T scanner, automatically optimized trajectories lead to significantly improved images, compared to standard low-pass, equispaced or variable density randomized designs. Insights into the nonlinear design optimization problem for MR imaging are given.

Keywords: Sparse Reconstruction, k-Space Optimization, Compressed Sensing, Experimental Design, Compressed Sensing, Bayesian Inference, Sub-Nyquist

1 Introduction

Modern MRI applications demand high imaging speed, especially when dynamic processes are observed, be it in cardiac examinations, angiography or fMRI studies at high temporal resolutions. This need for rapid scanning can be dealt with by alternative encoding strategies, making use of multiple receiver coils (1–3) in order to parallelize the imaging process to some extent.

While parallel MRI exploits redundancies between several receiver channels, imaging speed can also be increased by taking advantage of redundancies in the signal itself, which allows for reconstruction of the image from only a part of k-space in the first place. For example, k-space measurements show approximately Hermitian symmetry, which is exploited in partial Fourier acquisition techniques (4). Far beyond these simple symmetries, images form a statistically tightly constrained signal class. Fast, efficient digital image and video compression techniques are routinely used today, and the principles underlying them hold much promise for undersampled high resolution MRI reconstruction (5–8), if this process is understood in terms of nonlinear statistical estimation. These proposals are now known as *compressed sensing* or *sparse reconstruction*, since they exploit the statistical sparsity of images, a robust low-level characteristic, which leads to nonlinear, yet conservative and well-characterized interpolation behaviour (5). Compressed sensing is increasingly used for MRI problems, such as dynamic (9) and spectroscopic imaging (10), as well as for spiral (11) and radial undersampling (12, 13). Typically, scan time reductions by a factor of two or more can be achieved without losses in spatial resolution or sensitivity. Sparse statistics of images or image series originate from the structure of their pixel representations, an important instance of which is spatial or temporal redundancy, which has been used to speed up MRI acquisition (14–17).

Two problems arise in practical applications of compressed sensing: how to reconstruct an image from a fixed undersampling design, and how to choose the design in the first place. While a large amount of work was done for the former, we are not aware of much practically relevant progress for the latter. It is the undersampling design optimization problem for sparse MR image reconstruction that we focus on in this paper. Although there is substantial prior work on k-space optimization (18–20), this has been done for linear reconstruction, neglecting image sparsity. As we demonstrate here, it pays substantially to match the k-space trajectory to the sparse reconstruction technique. Nonlinear design optimization requires novel approaches. Established concepts such as the point spread function, taylored to linear reconstruction, do not capture the inherent dependence of sparse (nonlinear) estimation algorithms on the acquired signal. The latter cannot improve upon the Nyquist limit uniformly, but only for statistically sparse signals, and successful nonlinear k-space optimization has to take this dependence into account (the transform point

spread function proposed in (8) is signal-independent as well, and will not be useful to optimize k-space coverage for sparse reconstruction either). We phrase k-space optimization as a problem of experimental design, and propose an algorithm based on Bayesian inference, where statistical sparsity characteristics of images are incorporated by way of a prior distribution. The application of this procedure to high resolution MR images becomes feasible only with a novel inference algorithm we propose here. We apply our approach to the optimization of Cartesian and spiral trajectories, achieving a scan time reduction of a factor larger than two in either case, compared to Nyquist-spaced sampling. Our framework is generic and can in principle be applied to arbitrary trajectory classes, to multi-slice design optimization, and to designs with multiple receiver coils. We describe its computational and implementation requirements in detail. Properties of measurement designs for nonlinear sparse reconstruction have been evaluated empirically in (6) for Cartesian trajectories, and in (7, Sect. 6) for radial and spiral trajectories. They focus on non-convex image reconstruction and search for good designs by undirected random exploration, which is unlikely to cover the design space properly. In contrast, we employ the full Bayesian posterior in order to direct our search in a powerful and easily configurable manner. We are not aware of previous applications of (Bayesian) experimental design to this problem.

2 Theory

In this section, we develop a Bayesian experimental design framework for the optimization of k-space trajectories specifically for sparse nonlinear image reconstruction. Our method is based on a novel approximate Bayesian inference algorithm, which solves a convex optimization problem and can be scaled up to full high-resolution MR images.

2.1 Sparsity Statistics of MR Images. Convex Reconstruction

The Nyquist theorem fundamentally limits sampling designs without any assumptions on the signal. However, the vast majority of possible bitmaps do not constitute valid MR images, which are statistically tightly constrained. On a low level, images exhibit *sparsity*: coefficients in linear transform spaces have super-Gaussian distributions (peaked, heavy-tailed) (21), whose low entropies are mainly responsible for the high compression rates achieved by modern schemes such as JPEG. Sparsity is a robust property of non-synthetic images, coming from structure (edges, smooth areas, textures) which is not present in noise.

Sparsity can be used in order to reconstruct MR images from measurements far below the Nyquist limit. Let $\boldsymbol{u} \in \mathbb{C}^n$ represent the unknown MR image to be

reconstructed, where n is the number of pixels. MR measurements $\boldsymbol{y}$ are modelled as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{u} + \boldsymbol{\varepsilon} \in \mathbb{C}^m, \quad \Re(\boldsymbol{\varepsilon}),\, \Im(\boldsymbol{\varepsilon}) \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}),$$

where $\boldsymbol{\varepsilon}$ accounts for measurement errors, and $w = \Re(w) + \mathrm{i}\Im(w) \in \mathbb{C}$, $\Re(w), \Im(w) \in \mathbb{R}$. The design matrix $\boldsymbol{X} \in \mathbb{C}^{m \times n}$ contains Fourier filters at certain k-space points, and m is the number of k-space measurements taken. In standard linear reconstruction, we maximize the Gaussian likelihood $P(\boldsymbol{y}|\boldsymbol{u}) = N(\boldsymbol{X}\boldsymbol{u}, \sigma^2\boldsymbol{I})$ as a function of the bitmap $\boldsymbol{u}$, with no preference for sparse signals. This maximum likelihood estimator can be improved upon by taking signal class knowledge into account, in form of a *prior* probability distribution $P(\boldsymbol{u})$ over bitmaps. The prior is a preference weighting factor, unrelated to the measured data, which assigns high density to bitmaps exhibiting sparsity (such as MR images). Combining these terms by Bayes' rule, we have

$$P(\boldsymbol{u}|\boldsymbol{y}) \propto P(\boldsymbol{y}|\boldsymbol{u})P(\boldsymbol{u}).$$

$P(\boldsymbol{u}|\boldsymbol{y})$ is the Bayesian *posterior* distribution, the canonical combination of measurement data and prior knowledge by rules of probability. Both prior and posterior are distributions over bitmaps, representing our knowledge about the image before and after measurements have been obtained. In sparse reconstruction techniques, the posterior is optimized, instead of the likelihood alone. The most prominent algorithm, the MAP (maximum a posteriori) estimator, finds the mode of the posterior:

$$\hat{\boldsymbol{u}}_{\mathrm{MAP}} := \mathrm{argmin}_{\boldsymbol{u}}[-\log P(\boldsymbol{y}|\boldsymbol{u}) - \log P(\boldsymbol{u})]. \tag{1}$$

The super-Gaussian prior distribution $P(\boldsymbol{u})$ employed in this paper can be written as product of Laplace potentials, depending on linear projections of $\boldsymbol{u}$: multiscale wavelet coefficients and horizontal and vertical first derivatives. For this prior, the MAP estimator coincides with the method used in (8), apart from us placing additional Laplace potentials on the coefficients of $\Im(\boldsymbol{u})$ (13). Details are given in Appendix .1. The MAP reconstruction process, which is nonlinear due to the non-Gaussian prior $P(\boldsymbol{u})$, is illustrated in (8, Figure 2). As opposed to the maximum likelihood estimator, $\hat{\boldsymbol{u}}_{\mathrm{MAP}}$ cannot be found by a single linear system, but requires iterative computation. It is the unique minimizer of a *convex* criterion, and efficient MAP algorithms are available (22, 23). We refer to $P(\boldsymbol{y}, \boldsymbol{u})$ as *sparse linear model* (SLM), due to the linearity of measurements and the sparsity enforced by the prior.

2.2 Bayesian k-Space Optimization

Within a class of measurement designs $\boldsymbol{X}$ of equal cost, which of them leads to the most successful sparse MAP reconstruction of MR images? While this question has

been addressed satisfactorily for *linear* reconstruction, by the concept of point spread functions, we are not aware of a theory for the nonlinear sparse counterpart that is convincing in practice. Properties of nonlinear reconstruction are fundamentally signal-dependent, and to our knowledge, no theory at present captures the signal class of high-resolution MR images properly.

We develop a variant of Bayesian sequential experimental design (or Bayesian active learning) in this section, in order to optimize k-space sampling automatically from data, *specifically* for subsequent sparse MAP reconstruction. The key idea is to monitor the posterior distribution $P(\boldsymbol{u}|\boldsymbol{y})$, the Bayesian representation of remaining uncertainty in the image reconstruction, as the design $\boldsymbol{X}$ is sequentially extended by new trajectory interleaves. In each round, among many potential extension candidates, we select one which leads to the largest expected reduction in posterior uncertainty.

Sampling trajectories in most MRI sequences are composed of smooth elements, such as spiral or radial interleaves, or Cartesian phase encodes. Our design algorithm operates on a candidate set $\mathcal{C} = \{\boldsymbol{X}_*\}$ of such elements, and appends one element to the design $\boldsymbol{X}$ in each round. It is outlined in Algorithm 1. The design score (or criterion) is the *information gain*, quantifying the amount of reduction in posterior entropy due to the measurement of an additional phase encode $\boldsymbol{X}_*$. More precisely, it quantifies the difference in uncertainty between the present state of knowledge $P(\boldsymbol{u}|\boldsymbol{y})$ and the refined state $P(\boldsymbol{u}|\boldsymbol{y}, \tilde{\boldsymbol{y}}_*)$ after a novel measurement $\tilde{\boldsymbol{y}}_*$ at $\boldsymbol{X}_*$. A natural measure for the amount of uncertainty in a distribution $P(\boldsymbol{u})$ is the (differential) entropy $\mathrm{H}[P(\boldsymbol{u})] = -\int P(\boldsymbol{u}) \log P(\boldsymbol{u})\, d\boldsymbol{u}$ (24), based on which the information gain is defined as

$$\mathcal{S}(\boldsymbol{X}_*; P(\boldsymbol{u}|\boldsymbol{y})) := \mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y})] - \mathrm{E}_{P(\tilde{\boldsymbol{y}}_*|\boldsymbol{y})}\left[\mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y}, \tilde{\boldsymbol{y}}_*)]\right], \tag{2}$$

where $\mathrm{E}_{P(\boldsymbol{u})}[\dots]$ denotes $\int(\dots)P(\boldsymbol{u})\, d\boldsymbol{u}$. The expectation over the posterior $P(\tilde{\boldsymbol{y}}_*|\boldsymbol{y}) = \mathrm{E}_{P(\boldsymbol{u}|\boldsymbol{y})}[N(\tilde{\boldsymbol{y}}_*|\boldsymbol{X}_*\boldsymbol{u}, \sigma^2\boldsymbol{I})]$ is required, since the outcome $\tilde{\boldsymbol{y}}_*$ for a candidate $\boldsymbol{X}_*$ is unknown at scoring time. Neither the posterior $P(\boldsymbol{u}|\boldsymbol{y})$ nor the score values $\mathcal{S}(\boldsymbol{X}_*; P(\boldsymbol{u}|\boldsymbol{y}))$ can be computed in closed form, but have to be approximated by novel techniques, which are detailed in the following subsection.

Our algorithm provides a goal-directed way to optimize k-space sampling. In each round, only a single new real measurement is required, while the effective search space, the set of all combinations of candidates, has exponential size in the number of rounds. This characteristic sets it apart from blindly randomized approaches, which explore the search space in stochastic, non-adaptive patterns, and tend to use many more real measurements than rounds. In practice, our algorithmic scheme has to be adjusted to constraints coming from the MR scanner setup. We come back to this point in the Discussion. Our method is visualized in Figure 1.

Algorithm 1 Bayesian design optimization algorithm

Require: Candidate set $\mathcal{C}$ of elements (interleaves, phase encodes). Initial design $\boldsymbol{X}$, measurement $\boldsymbol{y}$, corresponding posterior $P(\boldsymbol{u}|\boldsymbol{y})$.

repeat

(1) Compute score values $\mathcal{S}(\boldsymbol{X}_*; P(\boldsymbol{u}|\boldsymbol{y}))$ for all candidate elements $\boldsymbol{X}_* \in \mathcal{C}$.

(2) Append winning candidate $\boldsymbol{X}_*$ to $\boldsymbol{X}$, and remove it from $\mathcal{C}$.

(3) Acquire measurement $\boldsymbol{y}_*$ corresponding to $\boldsymbol{X}_*$, append it to $\boldsymbol{y}$.

(4) Recompute novel posterior $P(\boldsymbol{u}|\boldsymbol{y})$.

until $\boldsymbol{X}$ has desired size and $\boldsymbol{u}$ desired quality

2.3 Variational Approximate Inference

In order to compute design score values $\mathcal{S}(\boldsymbol{X}_*; P(\boldsymbol{u}|\boldsymbol{y}))$, we have to integrate over the posterior $P(\boldsymbol{u}|\boldsymbol{y})$. These computations, referred to as Bayesian *inference*, cannot be done exactly in the case of sparse linear models. We propose a novel method for SLM approximate inference, which scales up to high-resolution MR images, while being accurate enough to successfully drive nonlinear design optimization. To the best of our knowledge, this regime could not be attained by previous SLM inference methods. The intractable posterior $P(\boldsymbol{u}|\boldsymbol{y})$ is fitted by a Gaussian distribution $Q(\boldsymbol{u}|\boldsymbol{y})$, with the aim of closely approximating the posterior mean and covariance matrix. The fitting amounts to a convex optimization problem with a unique solution, which is efficiently found by a novel iterative algorithm.

Our approach makes use of a *variational* relaxation, which has been used before (25–27). The posterior $P(\boldsymbol{u}|\boldsymbol{y})$ is fitted with the closest Gaussian distribution $Q(\boldsymbol{u}|\boldsymbol{y})$ from a large approximation family. Since integrations against Gaussian densities *are* tractable even in high dimensions, the replacement $P(\boldsymbol{u}|\boldsymbol{y}) \rightarrow Q(\boldsymbol{u}|\boldsymbol{y})$ allows for design score computations on a large scale. Our prior $P(\boldsymbol{u})$ is a product of super-Gaussian (Laplace) potentials, each of which can be tightly lower bounded by Gaussian functions of any variance (which amounts to the mathematical definition of super-Gaussianity; see Figure 2). We use this property in order to choose the approximation family, and to formulate the variational problem. For the former, we start with $P(\boldsymbol{u}|\boldsymbol{y})$, but replace each prior potential by a Gaussian lower bound centered at zero. The variances $\boldsymbol{\gamma} = (\gamma_i) \in \mathbb{R}_+^q$ of these replacements parametrize the Gaussian family members $Q(\boldsymbol{u}|\boldsymbol{y}; \boldsymbol{\gamma})$. For the variational criterion $\phi(\boldsymbol{\gamma})$, we apply the same replacement to the log partition function

$$\log P(\boldsymbol{y}) = \log \int P(\boldsymbol{y}|\boldsymbol{u}) P(\boldsymbol{u})\, d\boldsymbol{u}, \tag{3}$$

the approximation target in most variational inference methods (posterior moments, such as mean and covariance, are obtained as derivatives of $\log P(\boldsymbol{y})$) (28), leaving us with a lower bound $-\phi(\boldsymbol{\gamma})/2 \leq \log P(\boldsymbol{y})$, which can be evaluated as a

Gaussian integral. The larger the lower bound, the tighter is the fit of $Q(\boldsymbol{u}|\boldsymbol{y})$ to $P(\boldsymbol{u}|\boldsymbol{y})$: $2\phi(\boldsymbol{\gamma})+\log P(\boldsymbol{y})$ is a convex upper bound to the Kullback-Leibler divergence $\mathrm{D}[Q(\boldsymbol{u}|\boldsymbol{y}) \,\|\, P(\boldsymbol{u}|\boldsymbol{y})]$, a standard measure for the difference between two distributions (24).

We establish that the variational inference problem $\min_{\boldsymbol{\gamma}} \phi(\boldsymbol{\gamma})$ is convex: there is a single best Gaussian fit $Q(\boldsymbol{u}|\boldsymbol{y})$ to $P(\boldsymbol{u}|\boldsymbol{y})$. Moreover, we propose a novel algorithm to find the minimum point of ϕ orders of magnitude faster than previous approaches we know of, rapid enough to address the k-space optimization problem. Revisiting Algorithm 1, we obtain our method in practice by replacing $P(\boldsymbol{u}|\boldsymbol{y}) \rightarrow Q(\boldsymbol{u}|\boldsymbol{y})$, which is fitted before starting the design loop, and refitted to the extended posterior at the end of each round, in step (4). Details about our inference algorithm are given in Appendix .2. The optimization is reduced to calling primitives of numerical computing a moderate number of times: reweighted least squares estimation, and approximate eigendecomposition. While the former is routinely used for linear and nonlinear MRI reconstruction, the latter seems specific to the inference problem and is required in order to approximate posterior covariances. These are further reduced, by standard algorithms of numerical mathematics, to signal processing primitives such as fast Fourier transform (FFT) or nonequispaced fast Fourier transform (NFFT).

Once $P(\boldsymbol{u}|\boldsymbol{y})$ is replaced by its closest Gaussian fit $Q(\boldsymbol{u}|\boldsymbol{y})$, the design score (2) can be computed in practice (step (1) in Algorithm 1). However, k-space optimization comes with large candidate elements $\boldsymbol{X}_*$ (the spiral interleaves used in our study consist of 3216 k-space points), and if many candidates are to be scored in each round, a naive computation is too slow. We detail our score computation approach in Appendix .2, making use of approximate eigendecomposition once more.

3 Methods

We consider design problems for *Cartesian* and *spiral* sequences. In either case, we extract or interpolate measurements corresponding to desired trajectories from scanner data recorded on an equispaced grid (Magneton Trio scanner, Siemens Medical Solutions, Erlangen, Germany; turbo spin echo (TSE) sequence, 23 echos per excitation, train of 120° refocusing pulses, each phase encoded differently, $1 \times 1 \times 4\,\mathrm{mm}^3$; different echo times and orientations, see Figure 4). Reconstructions $\hat{\boldsymbol{u}}$ are validated by the L_2 distance $\|\boldsymbol{u}_{\mathrm{true}} - |\hat{\boldsymbol{u}}|\|$, $\boldsymbol{u}_{\mathrm{true}}$ being the absolute value of the complete data reconstruction. We use sparse MAP reconstruction in general (1), with code as used in (8), comparing against linear reconstruction (zero filling with density compensation) (29, ch. 13.2.4) for Cartesian undersampling.

3.1 Cartesian Trajectories

In the Cartesian setup, we select individual k-space lines from 256 equispaced candidates (with $d = 256$ samples per line), the complete dataset corresponding to a standard Nyquist-sampled image acquisition. Multiplications with $\boldsymbol{X}$, $\boldsymbol{X}_*$ correspond to equispaced discrete Fourier transforms, for which we use FFTW (Fastest Fourier Transform in the West; `www.fftw.org/`).

All designs compared here start with the 32 lines closest to the origin, which leaves 224 lines to choose from. Based on this low frequency data, we estimate a phase map and postmultiply $\boldsymbol{X}$ in order to correct for phase noise, as in (8). Phase mapping helps sparse reconstruction, and is vital for Bayesian design optimization (see Discussion). For the equispaced designs *eq*, the remaining space is covered with $N_{\text{shot}} - 32$ equispaced lines. The low-pass designs *ct* occupy lines densely from the center outwards. Random designs *rd* are drawn according to the heavy-tailed variable density used in (8) (we modified their density to accommodate the smaller central slab), which accounts for the nonuniform spectral distribution of (MR) images specifically ($1/f$ spectral decay). Lines are drawn without replacement. In accordance with (8), we noted that drawing lines *uniformly* at random results in poor reconstructions (not shown). Our Bayesian design optimization technique makes use of the remaining 224 lines as candidate set $\mathcal{C}$. The optimization is done on a single slice (TSE, TE=92ms, sagittal orientation; Figure 3, left), featuring many details, while we present test reconstruction results on a wide range of different data, unknown during design optimization.

3.2 Spiral Trajectories

Interleaved outgoing Archimedian spirals employ trajectories $\boldsymbol{k}(t) \propto \theta(t) e^{i2\pi[\theta(t)+\theta_0]}$, $\theta(0) = 0$, where the gradient $\boldsymbol{g}(t) \propto d\boldsymbol{k}/dt$ grows to maximum strength at the slew rate, then stays there (29, ch. 17.6). Sampling along an interleave (azimuthal direction) respects the Nyquist limit. The number of revolutions N_r per interleave, and the number of interleaves N_{shot} determine the radial spacing, with scan time proportional to N_{shot}. We use $N_r = 8$, resulting in 3216 samples per interleave. Radial Nyquist spacing is attained for $N_{\text{shot}} \geq 16$. Candidates are interleaves, parametrized by the offset angle: $\boldsymbol{X}_* = \boldsymbol{X}_*(\theta_0)$, with $d = 3216$ rows. Samples do not lie on a regular grid: non equispaced FFT is used to multiply with $\boldsymbol{X}$, $\boldsymbol{X}_*$ (NFFT with Kaiser-Bessel kernel (29, ch. 13.2); `www-user.tu-chemnitz.de/~potts/nfft`). Our experiments are idealized, in that spiral sampling is simulated by NFFT interpolation from data acquired on a grid.

We compare MAP reconstruction under a number of design choices: equispaced (*eq*), uniformly drawn at random (*rd*), and optimized (*op*). Angles lie in $[0, 2\pi)$ in the first, and in $[0, \pi)$ in the second setting. All designs contain $\theta_0 = 0$. In

addition, *eq* uses $\theta_0 = j(k\pi/N_{\text{shot}}),\ j = 1, \ldots, N_{\text{shot}} - 1$; for *rd*, we draw $N_{\text{shot}} - 1$ angles uniformly at random from $\mathcal{C} = (k\pi/256)[1:255]$, averaging results over ten repetitions; for *op*, we start from the single interleave $\theta_0 = 0$ and use the candidate set $\mathcal{C}$. Here, $k \in \{1, 2\}$, depending on the setting. For $k = 2$, setups with $N_{\text{shot}} = 8$ halve the scan time, compared to Nyquist spacing. Designs are optimized on a single slice (Figure 3, left), featuring many details.

In the first setting ($k = 2$), the near-Hermitian symmetry of data means that *eq* is at a disadvantage for even N_{shot}. In order to correct for this fact, and to test the relevance of $\boldsymbol{u}$ being close to real-valued (after phase mapping and subtraction), we restrict angles to $[0, \pi)$ in a second setting ($k = 1$). By interpolating non-Cartesian sampling, we ignore characteristic errors of spiral acquisition in practice, which may diminish the impact of our findings (see Discussion).

4 Results

4.1 Cartesian Sequences

Reconstruction error results are given in Figure 3 (tested on slice used for design optimization) and Figure 4 (tested on wide range of other data, unknown during design optimization). If nonlinear MAP reconstruction is used for undersampled reconstruction, the optimized designs clearly outperform all other choices, especially with fewer lines (the left end, 64 lines, is 1/4 of the Nyquist rate). Low-pass designs outperform variable density random designs with few lines, while the latter improves from about 1/2 the Nyquist rate. In contrast, if linear reconstruction is used (Figure 3, right), only low-pass designs lead to acceptable reconstructions.

Importantly, the dominating part of improvements of optimized over other designs considered here generalizes to data never seen during optimization, as shown in Figure 4. This is the case even for axial orientations, depicting details different from the single sagittal slice the design was optimized on. As seen in the right panel, the improvements are consistent across echo times, orientations, and subjects, and their size scales with the reconstruction difficulty of the test slice.

MAP reconstructions for Cartesian sagittal data (TSE, TE=88ms, unknown during design optimization) are shown in Figure 5, for axial data (TSE, TE=92ms) in Figure 6, comparing different designs of 64 lines (1/4 Nyquist; scan time reduction by factor of 4). The superior quality of reconstructions for the optimized design is evident.

4.2 Spiral Sequences

MAP reconstruction errors for spiral undersampling are given in Figure 7. The left column shows performance on the data the angles were optimized over, while

in the right column, we test generalization behaviour on a range of different data. The lower row corresponds to the first setting, with offset angles $\theta_0 \in [0, 2\pi)$. As expected, *eq* for even N_{shot} does poorly, due to the almost-Hermitian symmetry of the data, while performing comparably to *op* for odd N_{shot}. In the second setting ($\theta_0 \in [0, \pi)$, upper row), *eq* and *op* perform similarly from $N_{\text{shot}} = 7$, with *op* outperforming *eq* for smaller designs. In comparison, drawing offset angles at random leads to much worse MAP reconstructions in either setting. As for Cartesian undersampling, the performance on different datasets, unknown at optimization time, is comparable to the behaviour on the training set, except that *eq* does substantially worse on axial than on sagittal scans.

5 Discussion

We have highlighted the importance of k-space sampling optimization tailored specifically to novel nonlinear sparse reconstruction algorithms, and have proposed a Bayesian experimental design framework, scaled up to this application for the first time. Our experimental results for Cartesian undersampling show that sparse reconstruction quality depends strongly on the sampling design chosen, with phase encodes optimized by our Bayesian technique outperforming other commonly used undersampling schemes, such as low-pass or variable density random designs (8). With optimized sampling, high-quality reconstructions are obtained if only half of all lines are measured, and useful images can be reconstructed at 1/4 of the Nyquist rate (Figure 5, Figure 6). The behaviour of undersampling designs is very different for linear reconstruction, where only low-pass measurements lead to good reconstructions (Figure 3, right), indicating that linear design optimization concepts, such as the point spread function, play a diminished role for nonlinear reconstruction, and that sampling optimization has to be matched to the reconstruction modality. The improvement of optimized over other design choices is most pronounced for fewer number of lines acquired. Importantly, even though designs are optimized on a single slice of data, a large part of these improvements generalizes to different datasets in our study, featuring other slice positions, subjects, echo times, and even orientations (Figure 4). Our results indicate that Bayesian design optimization can be used offline, adjusting trajectories on data acquired under controlled circumstances, and final optimized designs can be used for future scans. Our framework embodies the idea of adaptive optimization. The sampling design is adjusted based on a representative dataset (called training set), and if adequate measures for complexity control are in place (Bayesian sparsity prior, proper representation of posterior mass, sequential scheme of uncovering information only if asked for), good performance on the training set (Figure 3) tends to imply good performance on independent test

sets (Figure 4), thus successful generalization to similar future tasks.

Our framework is not limited to Cartesian sampling, as demonstrated by our application to spiral k-space optimization. However, our findings are preliminary in this case: spiral sampling was interpolated from data acquired on a Cartesian grid, and only the offset angles of dense Archimedian interleaves were optimized (instead of considering variable-density spiral interleaves as well). In this setting, designs optimized by our technique show comparable performance to spacing offset angles equally, while a randomization of these angles performs much worse.

In Bayesian design optimization, statistical information is extracted from one or few representative images used during training and represented in the posterior distribution, which serves as oracle to steer further acquisitions along informative directions. Importantly, and confirmed in further experiments (not shown), it is essential to optimize the design on MRI data for real-world subjects, or controlled objects of similar statistical complexity; simple phantoms do not suffice. While the latter are useful to analyze *linear* reconstruction, they cannot play the same role for nonlinear sparse reconstruction. Modern theory proves that overly simple signals (such as piecewise constant phantoms) are reconstructed *perfectly* from undersampled measurements, almost independently of the design used for acquisition (30, 31). This advantage of sparse reconstruction *per se*, for almost any design, does not carry over to real-world images such as photographs (32) or clinical resolution MR images (our results here). The relevance of design optimization grows with the signal complexity, and is dominatingly present for MR images of diagnostically useful content and resolution.

Variable density phase encoding sampling does not perform well at 1/4 of the Nyquist rate (Figure 5, Figure 6), if the density of (8) is used. For a different density with lighter tails (more concentrated on low frequencies), reconstructions are better at that rate, but are significantly worse at rates approaching 1/2 or more (results not shown). In practice, this drawback can be alleviated by modifying the density as the number of encodes grows. From our experience, a second major problem with variable density design sampling comes from the *independent* nature of the process: the inherent variability of independent sampling leads to uncontrolled gaps in k-space, which tend to hurt image reconstruction substantially. Neither of these problematic aspects is highlighted in (8), or in much of the recent compressed sensing theory, where incoherence of a design is solely focussed on. A clear outcome from our experiments here is that while incoherence plays a role for nonlinear reconstruction, its benefits are easily outweighted by neglecting other design properties. Once design sampling distributions have to be modified with the number of encodes, and dependencies to previously drawn encodes have to be observed, the problem of choosing such a scheme is equivalent to the design optimization problem, for which

we propose a data-driven alternative to trial-and-error here, showing how to partly automatize a laborious process which in general has to repeated from scratch for every new configuration of scanner setup and available signal prior knowledge.

Further issues will have to be addressed in a fully practical application of our method. We extracted (or interpolated) undersampling trajectories from data acquired on a complete Cartesian grid, which may be realistic for Cartesian undersampling, but neglects practical inaccuracies specific to non-Cartesian trajectories. Moreover, in multi-echo sequences, the ordering of phase encodes matters. For an immobile training subject/object, our sequential method can be implemented by nested acquisitions: running a novel (partial) scan whenever $\boldsymbol{X}$ is extended by a new interleave, dropping the data acquired previously. With further attendance to implementation and commodity hardware parallelization, the time between these scans will be on the order of a minute. Gradient and transmit or receive coil imperfections (or sensitivities), as well as distortions from eddy currents, may imply constraints for the design, so that less candidates may be available in each round. Such adjustments to reality will be simplified by the inherent configurability of our Bayesian method, where likelihood and prior encode forward model and known signal properties.

The near-Hermitian symmetry of measurements is an important instance of prior knowledge, incorporated into our technique by placing sparsity potentials on the imaginary part $\Im(\boldsymbol{u})$. This leads to marked improvements for sparse reconstruction, and is *essential* for Bayesian k-space optimization to work well. In addition, phase mapping and subtraction is required. Phase contributions substantially weaken image sparsity statistics, thereby eroding the basis sparse reconstruction stands upon. In the presence of unusual phase errors, specialized phase mapping techniques should be used instead. In future work, we aim to integrate phase mapping into our framework.

In light of the absence of a conclusive nonlinear k-space sampling theory and the well-known complexity of nonlinear optimal design, our approach has to be seen in the context of other realizable strategies. Designs can optimized by blind (or heuristic) trial-and-error exploration (6), which in general is much more demanding in terms of human expert and MRI scan time than our approach. Well-founded approaches fall in two classes: artificially simplified problems are solved optimally, or adaptive optimization on representative *real* datasets is used. We have commented above on recent advances in the first class, for extremely sparse, unstructured signals (30, 31), but these results empirically seem to carry little relevance for real-world signals. Our method falls in the second class, as an instance of nonlinear sequential experimental design (33, 34), where real-world signals are addressed directly, and for which few practically relevant performance guarantees are available.

Our approach to design optimization is sequential, adapting measurements to largest remaining uncertainties in the reconstruction of a single image. While we established sound generalization behaviour on unseen data in our experiments, real-time MRI (9), (29, ch. 11.4) may especially benefit from our sequential, signal-focussed approach. While our algorithm at present does not attain the high frame rates required in these applications, algorithmic simplifications, combined with massively parallel digital computation, could allow our framework to be used in the future in order to provide advanced data analysis and decision support to an operator during a running MRI diagnosis.

In future work, we will apply our methodology to real non-Cartesian measurements. Moreover, we will address the trajectory design problem across multiple neighbouring slices, where sparsity statistics are modelled in three dimensions, and dependencies between measurements in different slices are represented by a Markovian extension of the sparse linear model. We also aim to explore design optimization problems in the context of parallel MR imaging, or for 3D imaging with ultra-short repetition times.

.1 Sparse Linear Model

The sparse linear model (SLM) we use throughout the paper, comes with a Gaussian likelihood $P(\boldsymbol{y}|\boldsymbol{u}) = N(\boldsymbol{X}\boldsymbol{u}, \sigma^2\boldsymbol{I})$ and a super-Gaussian image prior $P(\boldsymbol{u}) \propto \prod_{i=1}^{q} t_i(s_i)$. Here, $\boldsymbol{s} = \boldsymbol{B}\boldsymbol{u} \in \mathbb{C}^q$ consists of linear filter responses, with $\boldsymbol{B} \in \mathbb{R}^{q\times n}$: the image gradient (horizontal and vertical discrete first derivatives; also called total variation coefficients), and coefficients for an orthonormal multi-scale wavelet transform (Daubechies 4, recursion depth 6), a total of $q \approx 3 \cdot n$ *Laplace* potentials of the form

$$t_i(s_i) = e^{-(\tau_i/\sigma)|s_i|}, \quad \tau_i > 0.$$

The Laplace distribution stands out among sparsity-enforcing potentials, in that $-\log t_i(s_i) = (\tau_i/\sigma)|s_i|$ is convex, so that the MAP estimator $\hat{\boldsymbol{u}}_{\mathrm{MAP}}$ can be computed as a convex quadratic program (23). MAP estimation for this SLM exactly corresponds to the sparse reconstruction method of (8). In order to enforce the fact that $\boldsymbol{u}$ is close to real-valued, we make use of n additional Laplace potentials on $\Im(s_i)$, as in (13), but not in (8). Since $s_i \in \mathbb{C}$ is represented by $(\Re(s_i), \Im(s_i)) \in \mathbb{R}^2$ internally, this amounts to a simple extension of $\boldsymbol{B}$.

Scale parameters τ_i are shared among all potentials of the same kind, but we allow for different values in wavelet coefficient, total variation, and imaginary part potentials. While Bayesian inference is approximated over primary parameters $\boldsymbol{u}$, hyperparameters τ_i, σ^2 are estimated in general. In our experiments, we optimized them on data not used for comparisons, then fixed these values for all subsequent

sampling optimization and MAP reconstruction runs.

.2 Large Scale Variational Inference

Our variational inference algorithm falls in the class of concave-convex, or difference-of-convexprogramming methods. It is partly inspired by (35). The key idea is to replace ϕ by a surrogate upper bound $\phi_{\boldsymbol{z}}$, whose minimization is much simpler, but eventually leads to the same minimum. A more detailed exposition is given in (36, 37). The algorithm is summarized in Algorithm 2.

Super-Gaussian potentials can be tangentially lower bounded by Gaussian functions at any width: $t_i(s_i) = \max_{\gamma_i>0} e^{-(|s_i/\sigma|^2/\gamma_i+h_i(\gamma_i))/2}$, where $h_i(\gamma_i)$ parameterizes the height of the lower bound and depends on $t_i(s_i)$ (see Figure 2). If $h(\boldsymbol{\gamma}) := \sum_i h_i(\gamma_i)$, then $h(\boldsymbol{\gamma}) = (\boldsymbol{\tau}^2)^T\boldsymbol{\gamma}$, $\boldsymbol{\tau} = (\tau_i)$ for Laplace potentials. Plugging this into (3), we obtain a tractable lower bound to the log partition function. Let $\boldsymbol{\Gamma} = \mathrm{diag}\boldsymbol{\gamma}$, and assume for now that $\boldsymbol{B}^T\boldsymbol{\Gamma}^{-1}\boldsymbol{B}$ is nonsingular. If $Q(\boldsymbol{u}) := C^{-1}e^{-(2\sigma^2)^{-1}\boldsymbol{u}^H\boldsymbol{B}^T\boldsymbol{\Gamma}^{-1}\boldsymbol{B}\boldsymbol{u}}$, the approximating Gaussians are

$$Q(\boldsymbol{u}|\boldsymbol{y}) = N(\boldsymbol{u}_*, \sigma^2\boldsymbol{A}^{-1}) \propto P(\boldsymbol{y}|\boldsymbol{u})Q(\boldsymbol{u}),$$

$\boldsymbol{A} := \boldsymbol{X}^H\boldsymbol{X} + \boldsymbol{B}^T\boldsymbol{\Gamma}^{-1}\boldsymbol{B}$. Moreover, for Gaussians we have that

$$\int P(\boldsymbol{y}|\boldsymbol{u})Q(\boldsymbol{u})\,d\boldsymbol{u} = |2\pi\sigma^2\boldsymbol{A}^{-1}|^{1/2}\max_{\boldsymbol{u}} P(\boldsymbol{y}|\boldsymbol{u})Q(\boldsymbol{u}),$$

and some algebra leads to

$$\phi(\boldsymbol{\gamma}) := -2\log\int P(\boldsymbol{y}|\boldsymbol{u})Q(\boldsymbol{u})\,d\boldsymbol{u} = \log|\boldsymbol{A}| + h(\boldsymbol{\gamma})$$
$$+\min_{\boldsymbol{u}} R(\boldsymbol{u},\boldsymbol{\gamma}), \quad R := \sigma^{-2}\left(\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{u}\|^2 + \boldsymbol{s}^H\boldsymbol{\Gamma}^{-1}\boldsymbol{s}\right).$$

By a continuity argument, this equation holds for singular $\boldsymbol{B}^T\boldsymbol{\Gamma}^{-1}\boldsymbol{B}$ just as well. Now, $(\boldsymbol{u},\boldsymbol{\gamma}) \mapsto R(\boldsymbol{u},\boldsymbol{\gamma})$ is jointly convex, so that $\boldsymbol{\gamma} \mapsto \min_{\boldsymbol{u}} R$ is convex. Here, $x \mapsto y$ anonymously refers to a function mapping x to y. It is proved in (36) that $\boldsymbol{\gamma} \mapsto \log|\boldsymbol{A}|$ is convex as well. Finally, beyond Laplace sites used in the application here, it is shown in (36) that $h(\boldsymbol{\gamma})$ is convex iff all $\log t_i(s_i)$ are concave: for general super-Gaussian potentials $t_i(s_i)$, the variational inference problem $\min_{\boldsymbol{\gamma}}\phi$ is convex iff this is the case for MAP estimation. To our knowledge, no equivalent characterization has been given for any other variational relaxation applicable to sparse linear models.

A standard gradient-based optimization of $\phi(\boldsymbol{\gamma})$ is too expensive in general to be practical, and our second major technical contribution consists of a novel scalable class of solvers for $\min_{\boldsymbol{\gamma}}\phi$. The most problematic term is $\log|\boldsymbol{A}|$, introducing strong couplings into $\phi(\boldsymbol{\gamma})$, so that even $\nabla_{\boldsymbol{\gamma}}\phi$ is hard to compute. We provide a principled

way to fit ϕ by an upper bound $\phi_{\boldsymbol{z}}$, tangent at the present value $\boldsymbol{\gamma}$, so that $\phi_{\boldsymbol{z}}$ lacks this coupling term and is much easier to minimize than ϕ itself. In the resulting double loop (or upper bound) minimization algorithm, the major computational difficulty of minimizing ϕ has to be faced only at the few points where $\phi_{\boldsymbol{z}}$ is refitted to ϕ (outer loop updates), but not during the inner loop minimization of $\phi_{\boldsymbol{z}}$.

Since $\boldsymbol{\gamma}^{-1} \mapsto \log|\boldsymbol{A}|$ is concave, we have $\log|\boldsymbol{A}| = \min_{\{z_i>0\}} \boldsymbol{z}^T(\boldsymbol{\gamma}^{-1}) - g^*(\boldsymbol{z})$ by Legendre duality (38), so that $\phi(\boldsymbol{\gamma}) \le \min_{\boldsymbol{u}} \phi_{\boldsymbol{z}}(\boldsymbol{u}, \boldsymbol{\gamma})$ with

$$\phi_{\boldsymbol{z}}(\boldsymbol{u}, \boldsymbol{\gamma}) := \boldsymbol{z}^T(\boldsymbol{\gamma}^{-1}) + (\boldsymbol{\tau}^2)^T\boldsymbol{\gamma} + R(\boldsymbol{u}, \boldsymbol{\gamma}) - g^*(\boldsymbol{z}).$$

$\phi_{\boldsymbol{z}}$ is jointly convex. For fixed $\boldsymbol{u}$, the minimizer is $\gamma_i = \sqrt{z_i + (|s_i|/\sigma)^2}/\tau_i$. Plugging this in, we are left with

$$\min_{\boldsymbol{u}} \left[\sigma^{-2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{u}\|^2 + 2\sum_{i=1}^{q} \tau_i \sqrt{z_i + (|s_i|/\sigma)^2} \right], \tag{4}$$

a penalized least squares problem, which can be solved very efficiently by the iteratively reweighted least squares (IRLS) algorithm (39). Each IRLS iteration requires a single linear system to be solved, with matrices of the same size and form as $\boldsymbol{A}$: a problem well studied in linear MRI reconstruction. IRLS tends to converge after less than 30 iterations, forming the *inner loop* of our algorithm. The minimizer $\boldsymbol{u}_*$ coincides with the mean of $Q(\boldsymbol{u}|\boldsymbol{y})$. At IRLS convergence, $\boldsymbol{z}$ and $g^*(\boldsymbol{z})$ are recomputed, so that $\phi_{\boldsymbol{z}}$ is tangent to ϕ at the current value of $\boldsymbol{\gamma}$:

$$\begin{aligned} &\boldsymbol{z} \leftarrow \nabla_{\boldsymbol{\gamma}^{-1}} \log|\boldsymbol{A}| = \mathrm{diag}^{-1}(\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T) = (\mathrm{Var}_Q[s_i|\boldsymbol{y}]), \\ &g^*(\boldsymbol{z}) \leftarrow \boldsymbol{z}^T(\boldsymbol{\gamma}^{-1}) - \log|\boldsymbol{A}|. \end{aligned}$$

Our algorithm iterates between these *outer loop updates* and IRLS inner loop minimizations until convergence, which is typically attained after less than five outer loop updates. These latter computations, which are particular to approximate inference and not normally needed in reconstruction algorithms, are substantially harder to compute than least squares solutions. They can be approximated sufficiently well using the Lanczos eigensolver algorithm (40): $\boldsymbol{A} \approx \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^H$, $\boldsymbol{Q} \in \mathbb{C}^{n\times k}$ unitary, $k \ll n$, where the largest and smallest eigenvalues and eigenvectors of $\boldsymbol{A}$ feature in the diagonal $\boldsymbol{\Lambda}$ and $\boldsymbol{Q}$. Then,

$$\boldsymbol{z} \approx \mathrm{diag}^{-1}(\boldsymbol{B}\boldsymbol{Q}\boldsymbol{\Lambda}^{-1}\boldsymbol{Q}^H\boldsymbol{B}^T),\ \log|\boldsymbol{A}| \approx \log|\boldsymbol{\Lambda}|.$$

While this approximation is not very accurate uniformly over the z_i (they are systematically underestimated), it extracts enough information from $\boldsymbol{A}$ in order to drive our inference algorithm. This notion is discussed in more detail in (36). Intuitively,

the Lanczos algorithm extracts the dominating modes of $Q(\boldsymbol{u}|\boldsymbol{y})$ covariance (largest eigenvectors of $\boldsymbol{A}^{-1}$), much like in principal components analysis, and even for moderate k, these provide a sufficient summary of dependencies. The Lanczos method is related to the linear conjugate gradients (LCG) algorithm used to solve least squares systems, but requires $O(n\,k)$ memory and a computational overhead which grows with k. The rationale of our double loop algorithm is that, by bounding the $\log|\boldsymbol{A}|$ coupling term, this difficult covariance computation has to be done very few times only. Both LCG and Lanczos reduce much of their effort to repeated matrix-vector multiplications with $\boldsymbol{X}^H\boldsymbol{X}$ (fast Fourier transform, or NFFT). Our algorithm is summarized in Algorithm 2.

Algorithm 2 Double loop variational inference algorithm

Require: Data $\boldsymbol{X}$, $\boldsymbol{y}$.
  repeat
    if inference started from scratch **then**
      Initialize $\boldsymbol{z} \leftarrow 0.05 \cdot \boldsymbol{1}$, $g^*(\boldsymbol{z}) \leftarrow 0$, $\boldsymbol{u} \leftarrow \boldsymbol{0}$.
    else
      Outer loop update:
      Approximate eigendecomposition (Lanczos, k steps): $\boldsymbol{A} \approx \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^H$, $\boldsymbol{Q} = [\boldsymbol{q}_j] \in \mathbb{C}^{n\times k}$ unitary.

$$\boldsymbol{z} \leftarrow \sum_{j=1}^{k} \lambda_j^{-1} |\boldsymbol{B}\boldsymbol{q}_j|^2, \quad g^*(\boldsymbol{z}) \leftarrow \boldsymbol{z}^T(\boldsymbol{\gamma}^{-1}) - \sum_{j=1}^{k} \log\lambda_j.$$

      Initialize $\boldsymbol{u} \leftarrow \boldsymbol{u}_*$ (previous minimizer).
    end if
    Inner loop: Minimize (4) by IRLS algorithm, until minimizer $\boldsymbol{u}_*$ converges
    Update $\gamma_i \leftarrow (z_i + (|s_{*,i}|/\sigma)^2)^{1/2}/\tau_i$, $\boldsymbol{s}_* = \boldsymbol{B}\boldsymbol{u}_*$.
  until outer loop converged

Once $Q(\boldsymbol{u}|\boldsymbol{y})$ is fitted, design scores $\mathcal{S}(\boldsymbol{X}_*; Q(\boldsymbol{u}|\boldsymbol{y}))$ are computed by noting that $\mathrm{H}[Q(\boldsymbol{u}|\boldsymbol{y})] = \frac{1}{2}\log|2\pi e\sigma^2\boldsymbol{A}^{-1}|$, so that $\mathcal{S}(\boldsymbol{X}_*; Q(\boldsymbol{u}|\boldsymbol{y})) = \log|\boldsymbol{I} + \boldsymbol{X}_*\boldsymbol{A}^{-1}\boldsymbol{X}_*^H|$. Here, we approximate $P(\boldsymbol{u}|\boldsymbol{y}, \tilde{\boldsymbol{y}}_*)$ by $\propto Q(\boldsymbol{u}|\boldsymbol{y})P(\tilde{\boldsymbol{y}}_*|\boldsymbol{u})$ *without* refitting the variational parameters $\boldsymbol{\gamma}$. If $\boldsymbol{X}_* \in \mathbb{C}^{n\times d}$, $\mathcal{S}(\boldsymbol{X}_*)$ could be computed by solving d linear systems, but this is too slow to be useful. Instead, we use the Lanczos approximate eigendecomposition once more: $\log|\boldsymbol{I} + \boldsymbol{X}_*\boldsymbol{A}^{-1}\boldsymbol{X}_*^H| \approx \log|\boldsymbol{I} + \boldsymbol{V}_*^H\boldsymbol{V}_*|$, $\boldsymbol{V}_* := \boldsymbol{\Lambda}^{-1/2}\boldsymbol{Q}^H\boldsymbol{X}_*^H \in \mathbb{C}^{k\times d}$. If $k < d$, we compute $\log|\boldsymbol{I} + \boldsymbol{V}_*\boldsymbol{V}_*^H|$ instead. This approximation allows to score many large candidates in each round. Moreover, the score computation can readily be parallelized across different machines. We compared approximate score values to true ones, on 64×64 images where the latter can be computed. While the true values were strongly underestimated in general (even the largest ones), the peaks of the score curves were traced correctly by the approximations, and the maximizers of the approximate curves fell within dominating

peaks of the exact score.

Approximate inference is used at different points in Algorithm 1: in the initial phase before the design loop, and at the end of each round. In our experiments, we used 5 outer loop steps in the initial phase, and a single outer loop step between design extensions. We ran up to 30 inner loop IRLS steps, with up to 750 LCG iterations for each linear system (they often converged much faster). To save time, we partitioned the IRLS steps in categories "sloppy" and "convergence". Sloppy steps use 150 LCG iterations only, preceding convergence steps. The Lanczos algorithm was run for $k = 750$ iterations in general.

References

1. Sodickson DK, Manning WJ. Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiofrequency coil arrays. Magn Reson Med 1997;38:591–603.

2. Pruessmann K, Weiger M, Scheidegger M, Boesiger P. SENSE: Sensitivity encoding for fast MRI. Magn Reson Med 1999;42:952–962.

3. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A. Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn Reson Med 2002;47:1202–10.

4. McGibney G, Smith MR, Nichols ST, Crawley A. Quantitative evaluation of several partial Fourier reconstruction algorithms used in MRI. Magn Reson Med 1993;30:51–9.

5. Weaver J, Xu Y, Healy D, Cromwell L. Filtering noise from images with wavelet transforms. Magn Reson Med 1991;21:288–295.

6. Marseille G, de Beer R, Fuderer M, Mehlkopf A, Ormondt Dv. Nonuniform phase-encode distributions for MRI scan time reduction. J Magn Reson B 1996; 111:70–75.

7. Wajer F. Non-Cartesian MRI scan time reduction through sparse sampling. Ph.D. thesis, Delft University of Technology 2001.

8. Lustig M, Donoho D, Pauly J. Sparse MRI: The application of compressed sensing for rapid MR imaging. Magn Reson Med 2007;85:1182–1195.

9. Gamper U, Boesinger P, Kozerke S. Compressed sensing in dynamic MRI. Magn Reson Med 2008;59:365–373.

10. Hu S, Lustig M, Chen AP, Crane J, Kerr A, Kelley DA, Hurd R, Kurhanewicz J, Nelson SJ, Pauly JM, Vigneron DB. Compressed sensing for resolution enhancement of hyperpolarized 13C flyback 3D-MRSI. J Magn Reson 2008;192:258–64.

11. Santos J, Cunningham C, Lustig M, Hargreaves B, Hu B, Nishimura D, Pauli J. Single breath-hold whole-heart MRA using variable-density spirals at 3T. Magn Reson Med 2006;55:371–379.

12. Ye J, Tak S, Han Y, Park H. Projection reconstruction MR imaging using FOCUSS. Magn Reson Med 2007;57:764–775.

13. Block K, Uecker M, Frahm J. Undersampled radial MRI with multiple coils: Iterative image reconstruction with a total variation constraint. Magn Reson Med 2007;57:1086–1098.

14. Korosec F, Frayne R, Grist T, Mistretta C. Time-resolved contrast-enhanced 3D MR angiography. Magn Reson Med 1996;36:345–351.

15. Madore B, Glover G, Pelc N. Unalising by Fourier-encoding the overlaps using the temporal dimension (UNFOLD), applied to cardiac imaging and fMRI. Magn Reson Med 1999;42:813–828.

16. Tsao J, Boesinger P, Pruessmann K. k-t BLAST and k-t SENSE: Dynamic MRI with high frame rate exploting spatiotemporal correlations. Magn Reson Med 2003;50:1031–1042.

17. Mistretta C, Wieben O, Velikina J, Block W, Perry J, Wu Y, Johnson K, Wu Y. Highly constrained backprojection for time-resolved MRI. Magn Reson Med 2006;55:30–40.

18. Greiser A, von Kienlin M. Efficient k-space sampling by density-weighted phase-encoding. Magn Reson Med 2003;50:1266–75.

19. Von Kienlin M, Mejia R. Spectral localization with optimal point spread function. J Magn Reson 1991;94:268–287.

20. Spielman DM, Pauly JM, Meyer CH. Magnetic resonance fluoroscopy using spirals with variable sampling densities. Magn Reson Med 1995;34:388–94.

21. Simoncelli E. Modeling the joint statistics of images in the Wavelet domain. In Proceedings 44th SPIE 1999. 188–195.

22. Chen S, Donoho D, Saunders M. Atomic decomposition by basis pursuit. SIAM J Comp 1999;20:33–61.

23. Tibshirani R. Regression shrinkage and selection via the Lasso. J Roy Stat Soc B 1996;58:267–288.

24. Cover T, Thomas J. Elements of Information Theory. John Wiley & Sons, 1st edition 1991.

25. Girolami M. A variational method for learning sparse and overcomplete representations. N Comp 2001;13:2517–2532.

26. Palmer J, Wipf D, Kreutz-Delgado K, Rao B. Variational EM algorithms for non-Gaussian latent variable models. In Advances in NIPS 2006;18.

27. Jaakkola T. Variational methods for inference and estimation in graphical models. Ph.D. thesis, MIT 1997.

28. Jordan MI, ed. Learning in Graphical Models. Kluwer 1997.

29. Bernstein M, King K, Zhou X. Handbook of MRI Pulse Sequences. Elsevier Academic Press, 1st edition 2004.

30. Candès E, Romberg J, Tao T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Trans Inf Theo 2006;52:489–509.

31. Donoho D. Compressed sensing. IEEE Trans Inf Theo 2006;52:1289–1306.

32. Seeger M, Nickisch H. Compressed sensing and Bayesian experimental design. In Int. Conf. Mach. Learn. 2008;25.

33. Chaloner K, Verdinelli I. Bayesian experimental design: A review. Statistical Science 1995;10:273–304.

34. Fedorov V. Theory of Optimal Experiments. Academic Press 1972.

35. Wipf D, Nagarajan S. A new view of automatic relevance determination. In Advances in NIPS 2008;20. 1625–1632.

36. Seeger M, Nickisch H. Large scale variational inference and experimental design for sparse generalized linear models. Technical Report 175, MPI Biological Cybernetics, Tübingen 2008.

37. Seeger M, Nickisch H, Pohmann R, Schölkopf B. Bayesian experimental design of magnetic resonance imaging sequences. In Advances in NIPS 2009;21. 1441–1448.

38. Boyd S, Vandenberghe L. Convex Optimization. Cambridge University Press 2002.

39. Green P. Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. J Roy Stat Soc B 1984;46:149–192.

40. Schneider M, Willsky A. Krylov subspace estimation. SIAM J Comp 2001; 22:1840–1864.

List of Figures

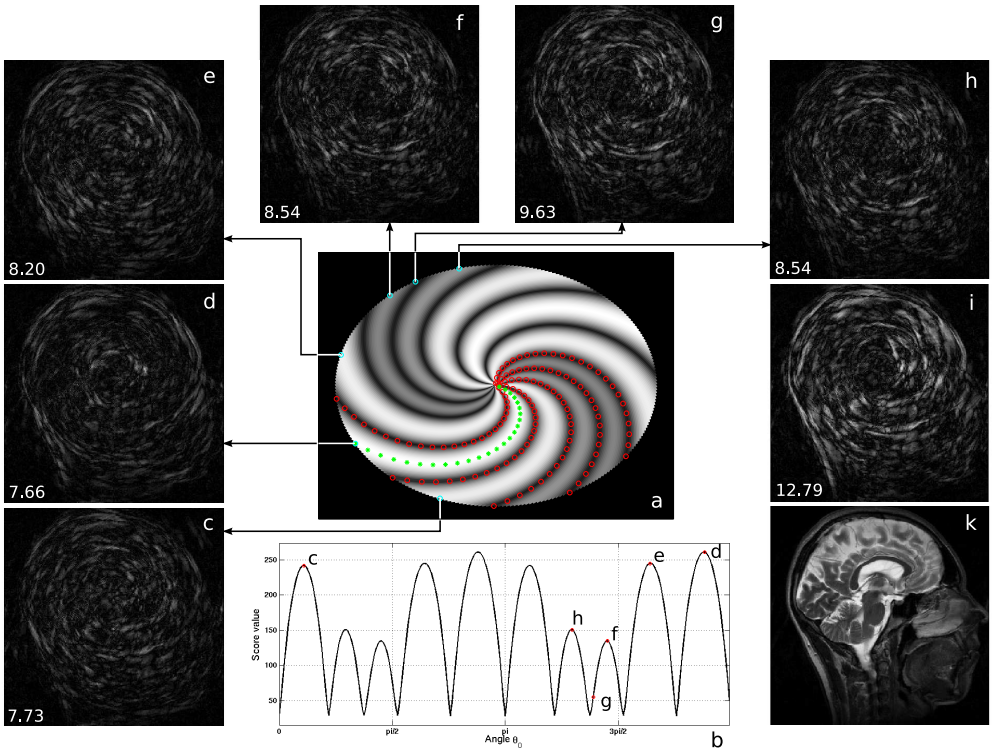

Figure 1: Bayesian experimental design on sagittal head scan data (see Methods section) for spiral sequences. Scoring round 5 → 6 interleaves. (a–b): Score values $\mathcal{S}(\boldsymbol{X}_*; Q(\boldsymbol{u}|\boldsymbol{y}))$ for 256 candidates $\theta_0 = k\,2\pi/256$, $k = 0:255$. (i): MAP reconstruction from $\boldsymbol{X}$ alone (5 arms). (c–h): MAP reconstructions from different design extensions $\boldsymbol{X} \cup \boldsymbol{X}_*$ (6 arms). Shown are residuals $|\boldsymbol{u}_* - \boldsymbol{u}_{\text{true}}|$ for reconstructions $\boldsymbol{u}_*$, L_2 error lower left. Top scorer (d) gives best reconstruction after extension, due to most information gained. Nontrivial score curve witnesses signal dependence of design optimization problem.

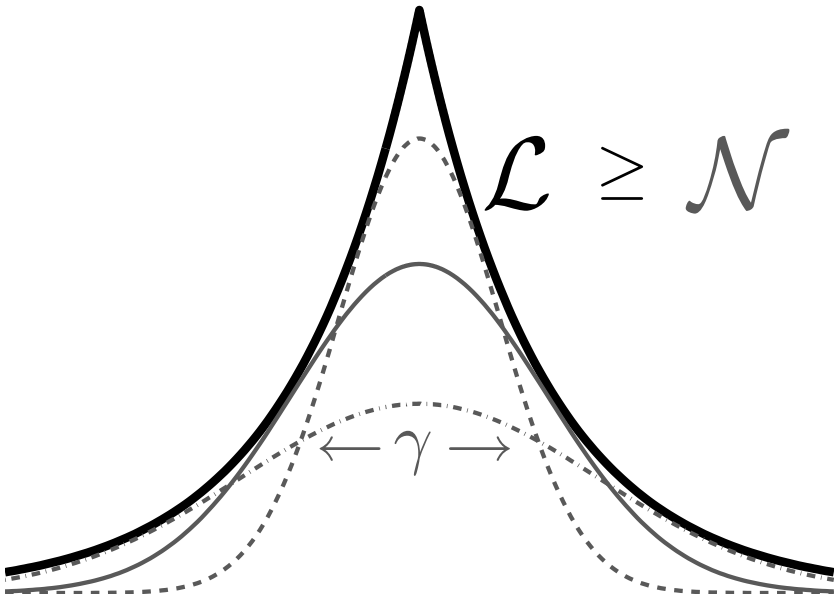

Figure 2: Super-Gaussian distributions (here: Laplace distributions, see Appendix .1) admit tight Gaussian-form lower bounds of any variance γ.

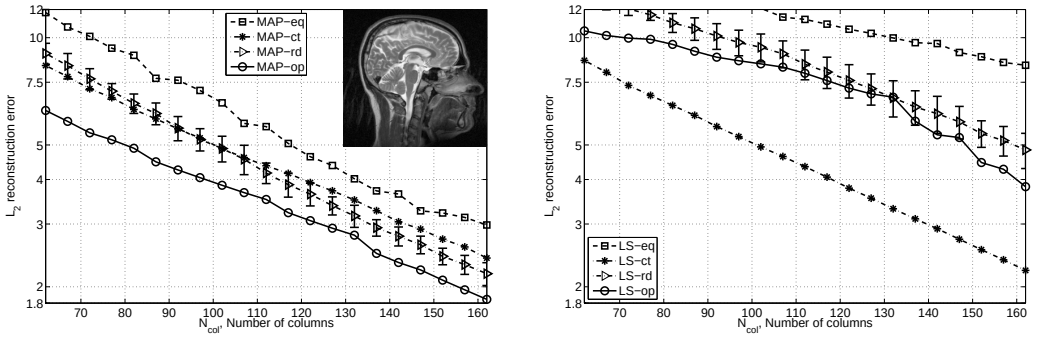

Figure 3: Results for Cartesian undersampling, on sagittal slice (TSE, TE=92ms). All designs contain 32 central lines. Equispaced [eq]; low-pass [ct]; random with variable density [rd]; optimized by our Bayesian technique [op], on same slice. Shown are L_2 distances to $\boldsymbol{u}_{\mathrm{true}}$. Left: Nonlinear (MAP) reconstruction. Right: Linear (ZFDC) reconstruction.

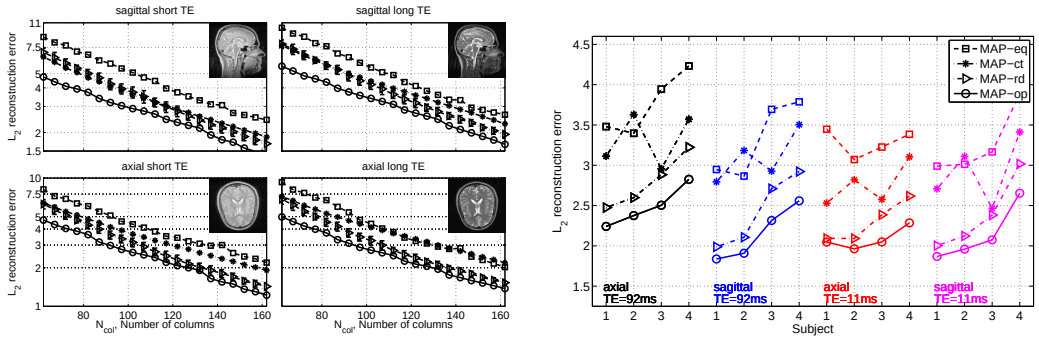

Figure 4: Results for Cartesian undersampling, on range of data unknown during design optimization: TSE scans, different echo times (TE=11ms, TE=92ms) and orientations (sagittal, axial). Design choices as in Figure 3. Shown are L_2 distances to $\boldsymbol{u}_{\text{true}}$, averaged over 5 slices and 4 different subjects. Left: Reconstruction test errors for different datasets (echo time, orientation). Error bars for variable density random [rd] w.r.t. ten repetitions. Right: Reconstruction test errors, averaged over 5 slices, for designs of 127 lines.

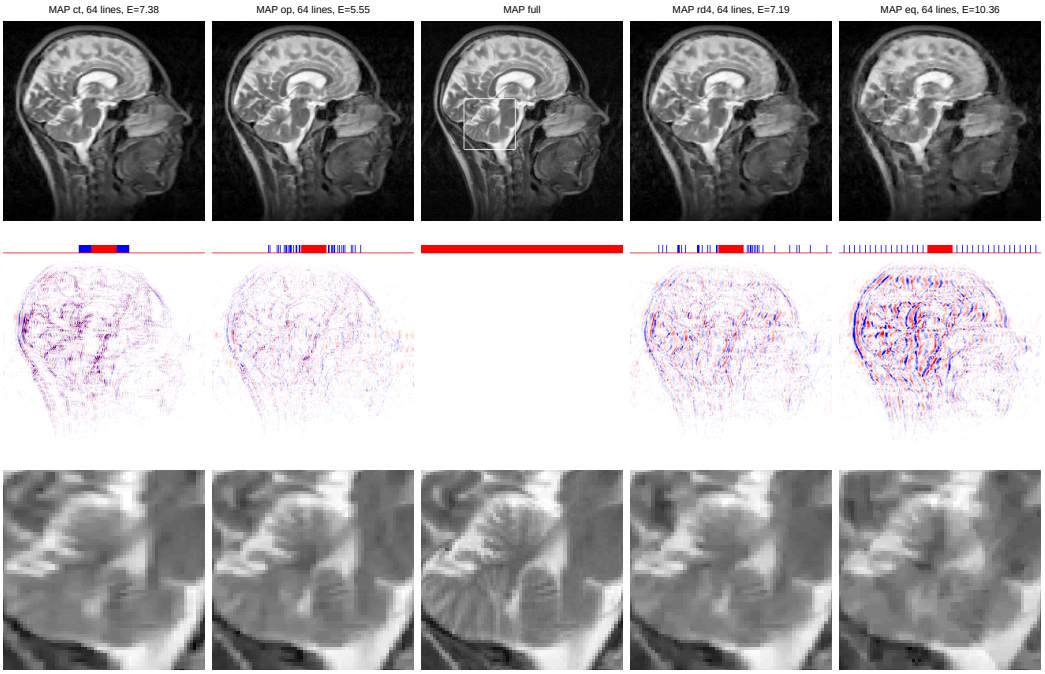

Figure 5: MAP reconstructions for Cartesian undersampling, sagittal data (TSE, TE=88ms, unknown during design optimization), at $N_{\text{shot}} = 64$ phase encodes (red: 32 initial center lines; blue: 32 additional encodes according to design choices). Upper row: Full images. White window: Location of blow-up. Middle row: Residuals (difference to $\boldsymbol{u}_{\text{true}}$), location of phase encodes (k-space columns). Lower row: Blow-ups.
MAP ct: Apparent lower resolution, fine structures smoothed out. MAP rd: Erroneous dark structure (upper left). MAP op: Satisfying level of details at 1/4 of Nyquist rate, considerably more detail and less blurring than for the other undersampled designs.

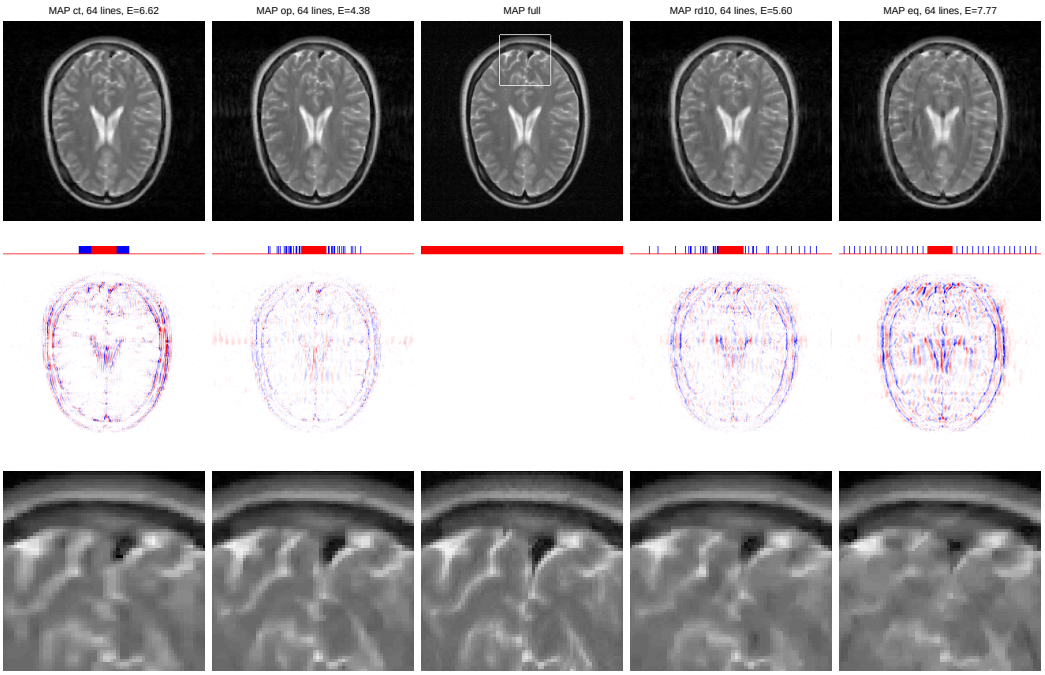

Figure 6: MAP reconstructions for Cartesian undersampling, axial data (TSE, TE=11ms, unknown during design optimization), at $N_{\text{shot}} = 64$ phase encodes (red: 32 initial center lines; blue: 32 additional encodes according to design choices). Upper row: Full images. White window: Location of blow-up. Middle row: Residuals (difference to $\boldsymbol{u}_{\text{true}}$), location of phase encodes (k-space columns). Lower row: Blow-ups.
MAP ct: Apparent lower resolution than MAP rd, MAP op. Both MAP ct and MAP rd have tendency to fill in dark area. MAP op retains high contrast there.

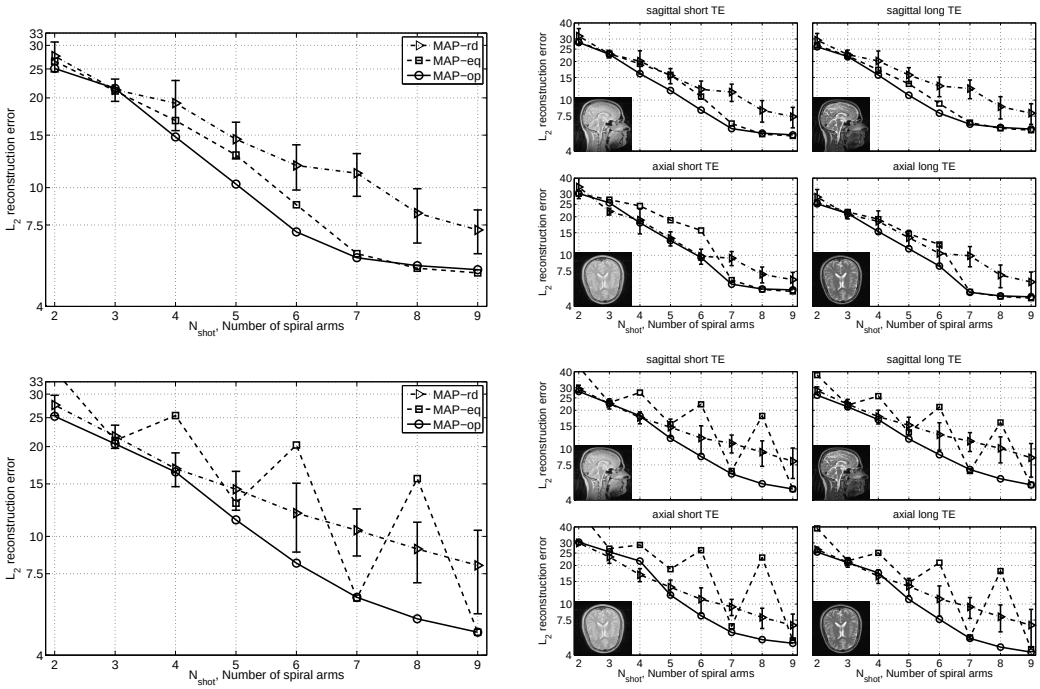

Figure 7: Results for MAP reconstruction, spiral undersampling of offset angles θ_0. Left column: Reconstruction errors on sagittal slice (see Figure 3 left), on which *op* is optimized. Right column: Reconstruction errors on different data (averaged over 5 slices, 4 subjects each, see Figure 4). Upper row: Offset angles from $[0, \pi)$. Lower row: Offset angles from $[0, 2\pi)$. Design choices: Equispaced [eq]; uniform at random [rd] (averaged over 10 repetitions); optimized by our Bayesian technique [op].